Production of Complex Human Glycoproteins in Yeast

Stephen R. Hamilton,* Piotr Bobrowicz,* Beata Bobrowicz,* Robert C. Davidson,* Huijuan Li,* Teresa Mitchell,* Juergen H. Nett,* Sebastian Rausch,* Terrance A. Stadheim,* Harry Wischnewski,* Stefan Wildt,* Tillman U. Gerngross†

We report the humanization of the glycosylation pathway in the yeast *Pichia pastoris* to secrete a human glycoprotein with uniform complex N-glycosylation. The process involved eliminating endogenous yeast glycosylation pathways, while properly localizing five active eukaryotic proteins, including mannosidases I and II, *N*-acetylglucosaminyl transferases I and II, and uridine 5′-diphosphate (UDP)–*N*-acetylglucosamine transporter. Targeted localization of the enzymes enabled the generation of a synthetic in vivo glycosylation pathway, which produced the complex human *N*-glycan *N*-acetylglucosamine$_2$–mannose$_3$–*N*-acetylglucosamine$_2$ ($GlcNAc_2Man_3GlcNAc_2$). The ability to generate human glycoproteins with homogeneous *N*-glycan structures in a fungal host is a step toward producing therapeutic glycoproteins and could become a tool for elucidating the structure-function relation of glycoproteins.

The ability to produce recombinant human proteins has led to major advances in human health care and remains an active area of drug discovery. Some proteins such as insulin, which are not glycosylated in their native state, do not require glycosylation and can thus be expressed in prokaryotic hosts that lack a glycosylation machinery, such as *Escherichia coli*. Most therapeutic proteins, however, require the cotranslational addition of glycans to specific asparagine residues of the protein to ensure proper folding and subsequent stability in the human serum. For therapeutic use in humans, glycoproteins require human-like N-glycosylation. Mammalian cell lines that are able to replicate human-like glycoprotein processing have several drawbacks including low protein titers, long fermentation times, heterogeneous products, and ongoing viral containment issues.

Yeast and filamentous fungi are robust industrial fermentation organisms that can be grown to high cell density in chemically defined medium. However, glycoproteins derived from fungal expression systems contain nonhuman *N*-glycans of the high mannose type (Fig. 1), which are immunogenic in humans and thus of limited therapeutic value (*1*). Nevertheless, fungi and mammals share initial steps of protein N-glycosylation, which involve the following: (i) the site-specific transfer of $Glc_3Man_9GlcNAc_2$ from the luminal side of the endoplasmic reticulum (ER) to the de novo synthesized protein by an oligosaccharyltransferase complex, (ii) trimming by glucosidases I and II, and (iii) the removal of one specific terminal α-1,2-mannose residue by an ER-residing α-1,2-mannosidase. These sequential steps lead to the formation of $Man_8GlcNAc_2$ structures (isomer Man8B) (Fig. 1), the *N*-glycan found on most glycoproteins leaving the fungal or mammalian ER. After the export of predominantly $Man_8GlcNAc_2$-containing glycoproteins to the Golgi, *N*-glycan processing pathways diverge notably between mammals and yeast (*2*). The human Golgi contains several α-1,2-mannosidases (IA, IB, and IC), which remove mannose to yield $Man_5GlcNAc_2$, a precursor for complex *N*-glycan formation (Fig. 1).

In *Saccharomyces cerevisiae*, N-glycosylation involves the addition of numerous mannose sugars throughout the entire Golgi, which often leads to hypermannosylated *N*-glycans with more than 100 mannose residues. When this first α-1,6-mannose is added by 1,6-mannosyltransferase (Och1p), additional α-1,6-mannosyltransferases extend the α-1,6-chain, which then becomes the substrate for α-1,2- and α-1,3-mannosyltransferases, as well as phospho-mannosyltransferases that add yet more mannose sugars to the growing *N*-glycan structure (*3*).

Pichia pastoris is a methylotrophic yeast with a glycosylation machinery that is similar to that of *S. cerevisiae* and is frequently used for the expression of heterologous proteins (*4*). YJN201, a recombinant strain of *P. pastoris*, lacks endogenous Och1p activity, and contains three heterologous genes encoding α-1,2-mannosidase localized to the ER, UDP-GlcNAc transporter, and GlcNAc transferase I localized to the Golgi (Fig. 1) (*5*). Two of these proteins (MnsI and GnTI) are synthetic fusions between fungal type II membrane proteins and catalytic domains from *Caenorhabiditis elegans* and humans, respectively. YJN201 has the ability to secrete the kringle 3 domain of human plasminogen (K3) as a reporter protein, with *N*-glycans of the hybrid type (*5*). Here, we further humanize the N-glycosylation pathways in *P. pastoris*.

To identify a compatible combination of MnsII and GnTII, a two-stage approach was chosen. First, a combinatorial library of several MnsII catalytic domains was fused to a signal peptide library of more than 60 fungal type II membrane localization signals. The resulting 500 combinatorial fusion constructs were introduced into the *P. pastoris* strain YSH1, described in the subsequent paragraph, capable of producing the human precursor of complex glycosylation, $GlcNAcMan_5GlcNAc_2$ (Fig. 2B, compare Fig. 2A) on the reporter K3. Only a small subset of strains (<5%) were capable of quantitatively converting $GlcNAcMan_5GlcNAc_2$ to $GlcNAcMan_3GlcNAc_2$ (Fig. 2C). These strains were isolated and subsequently transformed with a combinatorial library of several hundred GnTII–leader peptide fusions. Screening for the presence of $GlcNAc_2Man_3GlcNAc_2$ allowed the isolation of strains that were able to secrete homogeneous complex glycan (Fig. 2D), as exemplified by strain YSH44.

Pichia pastoris YSH44 was engineered from BK64-1 (*6*), an *och1* deletion mutant secreting K3, a reporter protein with a single N-linked glycosylation site (*5*). After the introduction of the *Kluyveromyces lactis* uridine 5′-diphosphate (UDP)-GlcNAc transporter (*5*), the strain was further used for the introduction and screening for active α-1,2-mannosidase fusions. Among 608 chimera, the mouse MnsIA catalytic domain fused to the N-terminal localization peptide of the ER protein Sec12 from *S. cerevisiae* was selected because of its ability to produce primarily $Man_5GlcNAc_2$. Subsequent introduction of a human GnTI fusion library allowed the isolation of a strain (PBP-3) able to produce primarily $GlcNAcMan_5GlcNAc_2$ structures (Fig. 2B). The URA marker in PBP-3 was recovered, which produced the strain YSH1. Transformation of the latter strain with a library of MnsII catalytic domains fused to a library of localization peptides permitted the isolation of yeast strains producing $GlcNAcMan_3GlcNAc_2$ as the primary *N*-glycan, with some $GlcNAcMan_4GlcNAc_2$ and $GlcNAcMan_5GlcNAc_2$ also present (Fig. 2C). This strain, YSH37, contains a catalytic domain of the *Drosophila melanogaster* mannosidase II (*7*), fused to the N-terminal localization signal of the *S. cerevisiae* Golgi protein Mnn2. Finally, we generated YSH44 by introducing a GnTII–localization peptide library and screening for strains that are able to

Thayer School of Engineering and the Department of Biological Sciences, Dartmouth College, Hanover, NH 03755, USA.

*Present address: GlycoFi, Inc., 21 Lafayette Street, Suite 200, Lebanon, NH 03766, USA.
†To whom correspondence should be addressed. E-mail: tillman.gerngross@dartmouth.edu

produce K3 with uniform $GlcNAc_2Man_3$ $GlcNAc_2$ structures (Fig. 2D). Specifically, this strain contained the catalytic domain of GnTII from *Rattus norvegicus* and the leader peptide from *S. cerevisiae* protein Mnn2. As expected, the *N*-glycans from YSH44 were substrates for both hexosaminidase and β-1,4-galactosyltransferase (Fig. 2, E and F). Unlike the parent strain BK64-1, YSH44 no longer displayed flocculation behavior in culture and was able to produce similar amounts of recombinant protein when compared with the commercially available wild-type *P. pastoris* strain GS115 (Fig. 3). These strains both demonstrated a similar degree of occupancy on the single N-linked site of K3 (~100%) (Fig. 3).

Being able to produce *N*-glycans of high purity in vivo depends on the efficiency of individual processing steps throughout the secretory pathway and the absence of local competing reactions for the same substrate. For example, YSH37 shows the consequence of producing a glycan intermediate in vivo that can also serve as a substrate for competing pathways. Although $GlcNAcMan_3GlcNAc_2$ represented the main *N*-glycan structure produced by this strain, we found two additional structures with masses corresponding to $GlcNAcMan_4GlcNAc_2$ and $GlcNAcMan_5GlcNAc_2$. We performed detailed analyses by mass spectrometry, high-performance liquid chromatography, and enzymatic digestions to determine the structure of these contaminating glycans. The fourth and fifth mannose residues on the $GlcNAcMan_3GlcNAc_2$ core were sensitive to α-1,2-mannosidase digest, which led us to conclude that mannosidase II conversion was complete and that subsequent α-1,2-mannosyltransferase activity was involved.

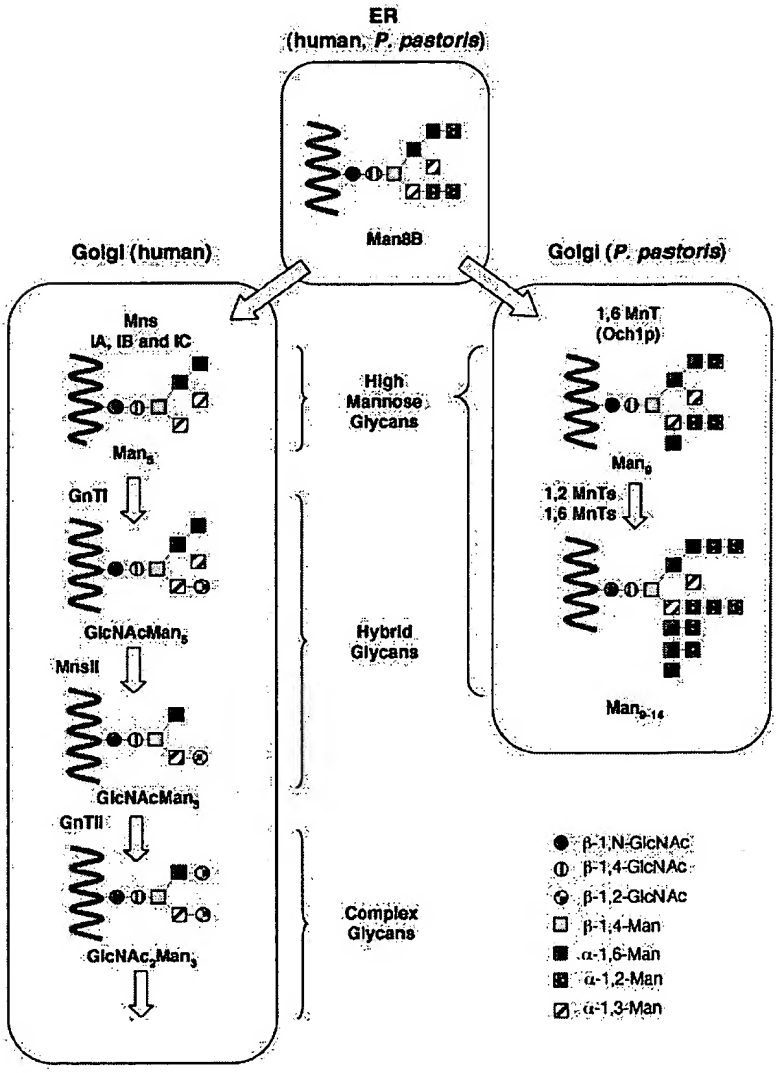

Fig. 1. N-linked glycosylation pathway in humans and in *P. pastoris*. Mns, α-1,2-mannosidase; MnsII, mannosidase II; GnT1, β-1,2-*N*-acetylglucosaminyltransferase I: GnTII,β-1,2-*N*-acetylglucosaminyltransferase II; MnT, mannosyltransferase.

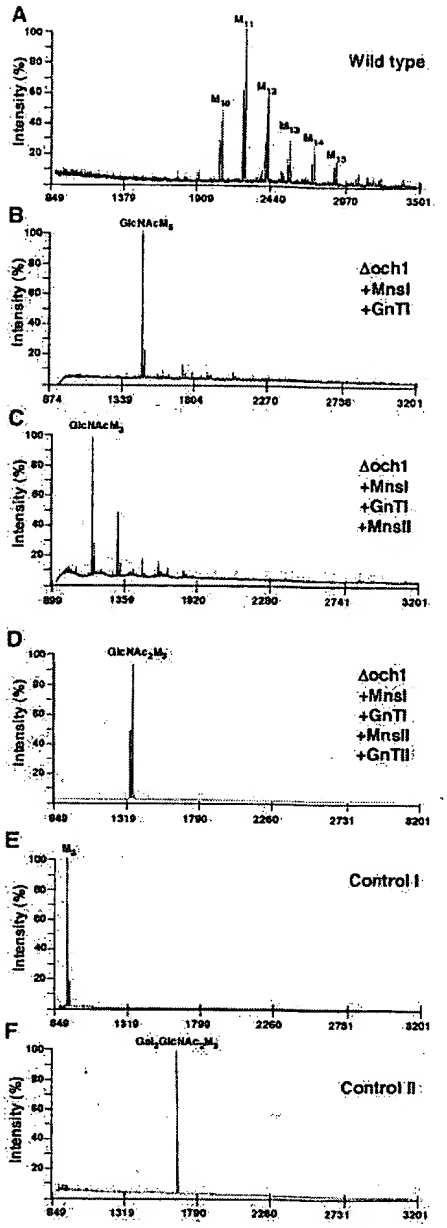

Fig. 2. Matrix-assisted laser desorption time-of-flight mass spectrometry (MALDI-TOF) spectra of N-linked glycans released from recombinant K3. K3 was produced in *P. pastoris* strains BK64 (*5*) [GS115 wild-type control (Invitrogen, Carlsbad, USA)], PBP-3, YSH37, and YSH44 and purified from culture supernatants by Ni-affinity chromatography. The glycans were released from K3 by peptide–*N*-glycosidase F treatment. The released N-linked glycans were analyzed by MALDI-TOF mass spectrometry (positive mode), typically appearing as the sodium or potassium adducts. (**A**) BK64, (**B**) PBP-3, (**C**) YSH37, (**D**) YSH44, (**E**) glycans from YSH44 after β-*N*-acetylhexosaminidase treatment, and (**F**) glycans from YSH44 after β-1,4-galactosyltransferase treatment. To simplify labeling of glycoforms, the two core GlcNAc residues, although present in all species, have been omitted. M, mannose.

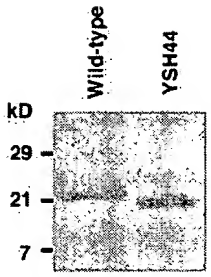

Fig. 3. Comparative expression levels of K3 in wild-type (BK64) (*5*) versus YSH44 (*6*) *P. pastoris*. Analysis of the reporter protein K3 by SDS–polyacrylamide gel electrophoresis (SDS-PAGE) (*6*).

This was further substantiated by the introduction of GnTII, which preempted the action of α-1,2-mannosyltransferase and led to the formation of uniform $GlcNAc_2Man_3GlcNAc_2$ structures (Fig. 2, D and E).

In the glycosylation process, multiple enzymes compete for the same transient glycan structures, which typically leads to heterogeneous mixtures of glycoforms. Here the fungal glycosylation pathway was reengineered to mirror the processing of human *N*-glycan structures. However, unlike the human pathway, which typically results in an array of glycoforms, this system yielded essentially homogeneous glycoforms.

To better understand the relation between N-glycosylation and protein function, generating uniform glycoproteins with specific *N*-glycan structures will prove to be of great utility (*8*). At present, the interpretation of these relations is compromised by the heterogeneity of the glycoform pools that can be generated from mammalian sources. These pools are typically created by expressing glycoproteins in specific glycosylation mutants or by ex vivo enrichment, such as lectin chromatography or enzymatic treatment. Moreover, even when a particular structure is identified, it is difficult to produce these structures at a commercial scale. Being able to engineer human glycosylation pathways into yeast strains able to express proteins with uniform *N*-glycan structures offers a more practical solution to this problem. When single proteins from a library of genetically engineered yeasts can be expressed and when each produces a defined and uniform glycoform, glycoprotein libraries can be generated to elucidate specific structure-function relations and to identify the most efficacious glycoform for a particular biological function. Once identified, a particular glycoform can be readily produced at industrial scale because of the rapid and well-established scale-up of yeast fermentations.

References and Notes

1. C. E. Ballou, *Methods Enzymol.* **185**, 440 (1990).
2. S. C. Hubbard, R. J. Ivatt, *Annu. Rev. Biochem.* **50**, 555 (1981).
3. N. Dean, *Biochim. Biophys. Acta* **1426**, 309 (1999).
4. J. M. Cregg, J. L. Cereghino, J. Shi, D. R. Higgins, *Mol. Biotechnol.* **16**, 23 (2000).
5. B. K. Choi *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 5022 (2003).
6. Materials and methods are available as supporting material on *Science* Online.
7. J. M. Foster, B. Yudkin, A. E. Lockyer, D. B. Roberts, *Gene* **154**, 183 (1995).
8. T. Shinkawa *et al.*, *J. Biol. Chem.* **278**, 3466 (2003).
9. We thank A. Prasad and N. Nutile-McMenemy for technical assistance and acknowledge the support of National Institute of Standards and Technology, Advance Technology Program grant 70NANB2H3046, and NIH Small Business Innovative Research grant 1R43GM66690-1.

Supporting Online Material

www.sciencemag.org/cgi/content/full/301/5637/1244/DC1

Materials and Methods

References

18 June 2003; accepted 18 July 2003

Representation of Action Sequence Boundaries by Macaque Prefrontal Cortical Neurons

Naotaka Fujii and Ann M. Graybiel*

Complex biological systems such as human language and the genetic code are characterized by explicit markers at the beginning and end of functional sequences. We report here that macaque prefrontal cortical neurons exhibit phasic peaks of spike activity that occur at the beginning and endpoint of sequential oculomotor saccade performance and have the properties of dynamic start- and end-state encoders accompanying responses to sequential actions. Sequence bounding may thus reflect a general mechanism for encoding biological information.

Speaking, hitting a fastball, and playing the flute are premier examples of behaviors requiring meticulous sequencing of actions. These are learned skills, but the ability to carry out sequences of movements is fundamental to most normal behaviors of humans and other animals (*1*). These sequential behaviors extend over time, so that temporal order as well as spatial order must be regulated. Many levels of the neuraxis have been implicated in the control of sequential behaviors (*2–6*). The prefrontal cortex, however, is a key region involved in the planning and automatization of the sequential behaviors that enable us to produce a smooth flow of actions or thoughts across time (*7–11*).

Implicit in the structure of such sequences is that they have a clear beginning and end, and explicit designations of start and end states have proven useful in models of sequential behavior (*12*). To test for such representations of start and end states, we recorded from multiple neurons in the prefrontal cortex and, for comparison, in the frontal eye field (FEF), as highly trained macaque monkeys performed up to seven sequential saccades (eye movements to redirect the line of sight) in blocks of trials (*13*) (Fig. 1A). Successive trials had a clear beginning (the illumination of a central red fixation point), a clear movement period (in which the monkey made saccades to successively presented red visual targets), and a final reward delivery period (in which the monkeys received juice or water for correct performance). In the standard task, different sequences were presented in pseudo-random order within trial blocks. We analyzed neuronal spike activity in relation to successive task events throughout the trials and during the intertrial intervals (*13*).

Neurons in the FEF responded selectively to saccades made in particular directions (*14, 15*), and some showed sequence selectivity. Many neurons in the prefrontal cortex exhibited similar properties, but in addition they responded to multiple phases of the task and often exhibited tonic firing during the entire trial or during the movement period (*13, 16*). Figure 1B illustrates typical prefrontal responses during the four-saccade task, with peaks of activity for each saccade during the movement period and a smaller peak after initiating fixation.

Nearly half of the task-related prefrontal neurons (295 of 658), but only rare FEF neurons (13 of 116), had an additional phasic peak in firing 270 to 280 ms after the sequence of saccades was completed. This "extra" peak occurred no matter how many saccades were made during the previous movement period (Fig. 1, B, E, and F) (fig. S1) (*13*). Even for a single saccade, there

Department of Brain and Cognitive Sciences and McGovern Institute for Brain Research, Massachusetts Institute of Technology, 45 Carleton Street, E25-618, Cambridge, MA 02139, USA.

*To whom correspondence should be addressed. E-mail: graybiel@mit.edu